Title

Biallelic mutations in the 3' exonuclease TOE1 cause pontocerebellar hypoplasia and uncover a role in snRNA processing.

Journal

Nature genetics, 49(3)

ISSN

1061-4036

Authors

Lardelli, Rea M
Schaffer, Ashleigh E
Eggens, Veerle RC
et al.

Publication Date

2017-03-01

DOI

10.1038/ng.3762

Peer reviewed

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2017 July 16.

Published in final edited form as:

Nat Genet. 2017 March ; 49(3): 457–464. doi:10.1038/ng.3762.

Biallelic Mutations in the 3' Exonuclease *TOE1* Cause Pontocerebellar Hypoplasia and Uncover a Role in snRNA Processing

A full list of authors and affiliations appears at the end of the article.

Abstract

Deadenylases are best known for degrading the poly(A) tail during mRNA decay. The deadenylase family has expanded throughout evolution and, in mammals, consists of 12 Mg^{2+}-dependent 3' end ribonucleases with mostly unknown substrate specificity[1]. Pontocerebellar hypoplasia type 7 (PCH7) is a unique recessive syndrome characterized by neurodegeneration with ambiguous genitalia[2] (MIM%614969). We studied 12 human families with PCH7, uncovering biallelic, loss of function mutations in *TOE1* (NC_000001.11), which encodes an unconventional deadenylase[3,4]. Toe1-morphant zebrafish displayed mid- and hind-brain degeneration, modeling PCH-like structural defects *in vivo*. Surprisingly, we found TOE1 associated with incompletely processed small nuclear (sn)RNAs of the spliceosome, which is responsible for pre-mRNA splicing. These pre-snRNAs contained 3' genome-encoded tails often followed by post-transcriptionally added adenosines. Human cells with reduced levels of TOE1 accumulated 3' end-extended pre-snRNAs, and immuno-isolated TOE1 complex was sufficient for 3' end maturation of snRNAs. Our findings reveal the cause of a neurodegenerative syndrome linked to snRNA maturation and uncover a key factor involved in processing of snRNA 3' ends.

**Addressed correspondence to: J.L-A. jlykkeandersen@ucsd.edu, F.B. f.baas@amc.uva.nl and J.G.G. jogleeson@rockefeller.edu.
*These authors contributed equally.

Author contributions

A.E.S., S.J.M., N.A., A.G.-G., L.D.H., and A.G. performed fibroblast, iPSC, NPC and knockout mouse experiments. R.M.L. generated stable cell lines and performed all snRNA/P experiments. V.R.E., N.A., E.S., D.M., R.M., A.W., A.E.S., E.D., and R.O.R. analyzed clinical and exome results. S.L.G., A.E.S., Z.S., B.R., A.G.-G., I.M.-V., and L.D.H. generated zebrafish data, and N.C. and D.T. provided resources for zebrafish experimentation. M.S.Z., N.F., W.B.D., L.S., S.M.B., E.M.V., J.H.D., L.D.M., H.K., U.A., M.L.F., L.W., D.C., C.F., M.K., K.A.A., D.M., N.M., A.O.Ç, K.B., H.P., M.-C.N., H.S., M.S., K.O., and K.M. conceived of genetic investigation. S.B.G., M.G. supported exome sequencing. E.L.V.N., S.S., T.L.S., and G.W.Y. supported RNA sequencing and computational analysis. E.J.B. performed mass spectrometry. R.L.M, A.E.S., J.L.-A. and J.G.G. wrote the manuscript. R.L.M., A.E.S., V.R.E., Z.S., N.C., D.T, G.W.Y., F.B., J.L.-A., and J.G.G. edited the manuscript. J.L.-A., F.B. and J.G.G. directed the project.

Competing financial interests

The authors declare no competing financial interests.

Accession codes. The exome sequencing data for all individuals consented for public release of data in this study have been deposited to the database of Genotypes and Phenotypes (dbGaP) under accession phs000288.v1.p1. RNA sequencing has been deposited into Gene Expression Omnibus (GEO) under accession GSE71536.

Online Resources.

NHLBI Exome Sequencing Project (ESP) http://evs.gs.washington.edu/EVS/
SeattleSeq http://snp.gs.washington.edu/SeattleSeqAnnotation137/
OMIM http://www.omim.org/
PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/http://genetics.bwh.harvard.edu/pph2/
Greater Middle Eastern Variome http://igm.ucsd.edu/gme/

Keywords

TOE1; CAF1Z; Pontocerebellar hypoplasia; snRNP; snRNA; deadenylation

Pontocerebellar neurodegeneration has its onset so early as to overlap with neurodevelopment and is thus alternatively referred to as pontocerebellar hypoplasia (PCH)[5]. PCH7 is characterized by neurological deterioration, atrophy/hypoplasia of the pons and cerebellum, muscular hypotonia, and breathing abnormalities, combined with hypogonadism[2]. This combination of rare conditions suggested a unique syndromic association due to a single gene mutation, but no locus or causative gene has been identified to date. We recruited 12 families meeting criteria for PCH7 including the index family that defined the condition (Fig. 1a) and confirmed that clinical features matched those published (Supplementary Table 1), including reduced pons/cerebellum parenchyma (Fig. 1b) ventriculomegaly, thin corpus callosum, and variable hypogonadism, ranging from absent gonads to ovarian and uterine remnants or atrophic and undescended testes (Supplementary Fig. 1a, b). All patients/families were enrolled in Institutional Review Board-approved protocols at referral institutions and provided consent for study. We performed whole exome sequencing in the proband and parents from Egyptian families 1275 and 1603. Aligning the genomic variants uncovered a p.E220K homozygous mutation in *TOE1* within a shared haplotype of 500 kb (Fig. 2a, Supplementary Fig. 2a, b), indicative of shared ancestry. All 10 additional families enrolled subsequently proved positive for biallelic mutations in *TOE1* (Fig. 2a, Supplementary Table 2). The variants were all predicted to impair expression of full length TOE1 or affect protein function[6] by altering well-conserved amino acids (Fig. 2b) and were not observed in our 4,000 in-house ethnically-matched exome database. We confirmed that each variant segregated according to a recessive mode of inheritance in all genetically informative members of each family, suggesting *TOE1* biallelic mutations underlie PCH7.

To test if TOE1 missense mutations were likely to interfere with protein function, we modeled them on the structure of the paralogous protein CNOT7 (Q9UIV1) (Supplementary Fig. 3a). Most variants were on surface residues, rather than within the RNA-binding cleft, suggesting they are not likely to directly affect deadenylase activity (Fig. 2c). We next established primary fibroblast cultures from patients carrying the homozygous p.E220K, p.F148Y and p.A103T mutations, as well as unaffected relatives (Family 1603, 4127 and 4128, respectively). Patient fibroblasts had reduced levels of TOE1 protein by Western blot (Fig. 2d, Supplementary Fig. 3b). Additionally, neural progenitor cell (NPC) lines derived from an affected patient of Family 1603 showed less TOE1 protein than an unaffected relative and control NPCs (Fig. 2e, Supplementary Fig. 3b, c). Together, these results indicate TOE1 amino acid substitutions negatively affect protein accumulation.

To verify *TOE1* mutations affect protein levels, we generated single-site integration T-REx-293 cell lines for tetracycline-regulated expression of siRNA-resistant wild type human TOE1 (WT) and p.E220K, p.F148Y, p.V173G, and p.A103T variants. When depleted of endogenous TOE1 and induced with a concentration of tetracycline that promoted accumulation of TOE1-WT to near endogenous levels, we observed reduced levels of mutant

compared with TOE1-WT, despite similar mRNA levels (Supplementary Fig. 3d, e). By contrast, a previously characterized catalytically inactive mutant of TOE1 (DE) accumulated to similar levels as WT (Supplementary Fig. 3e)[4]. These results are consistent with our findings from patient cell lines, and indicate the patient mutations are deleterious by reducing TOE1 levels.

Since *Toe1* is expressed broadly in human and mouse tissue (Supplementary Fig. 4a, b), we generated *Toe1* mutant mice to test for defects *in vivo*. Embryos with homozygous *Toe1* frameshift mutations showed uniform lethality prior to embryonic day (e) 11.5 (Supplementary Fig. 4c), demonstrating *Toe1* is required for mouse development. Because the patient mutations allow for partial expression of TOE1 protein, we next turned to morpholino (MO) knockdown in zebrafish, where dosage could be regulated, to create a PCH7 disease model (Supplementary Fig. 5a). Knockdown of the single *toe1* orthologue (NM_001256682.1) led to reproducible microcephaly, small eye, and curly tail phenotype in 90% of embryos by 48 hours post fertilization (hpf), which was rescued by co-injection of human *TOE1* mRNA but not the catalytically inactive DE mutant- or patient mutation-encoding mRNAs (Fig. 3a, Supplementary Fig. 6a). To visualize neurons, we performed whole mount immunofluorescence for the neuronal marker, HuC. Like patients, *toe1* MO injected fish showed structural defects of the developing midbrain, cerebellum and hindbrain (Fig. 3b). This phenotype was largely restored by co-injection of WT zebrafish or human *TOE1*, but not mutant mRNAs (Fig. 3b, Supplementary Fig. 5b, 6b). To determine if neuronal loss in the *toe1* MO injected zebrafish was due to cell death, we performed cleaved caspase-3 (Casp3) staining at 24 hpf. While zebrafish injected with Control MO showed few Casp3-positive cells, *toe1* MO-injected zebrafish showed a dramatic apoptotic response, consistent with neurodegeneration (Supplementary Fig. 5c). We conclude that reduced expression of TOE1 leads to neurodegeneration and PCH-like structural brain defects *in vivo*.

TOE1 was originally identified as a growth suppressor and direct target gene of Erg1, an immediate early transcription factor[3]. Alternatively called CAF1z due to homology with CAF1 deadenylases, we reported TOE1 as a 3’ to 5’ exonuclease with preference for adenosines *in vitro*[4]. Unlike characterized mRNA deadenylases, TOE1 concentrated in nuclear Cajal Bodies (CBs)[4], which are rich in enzymes that process RNAs not known to have poly(A) tails, suggesting novel substrates. More recently, TOE1 was shown to associate with spliceosomal proteins[7,8], which are known to localize to CBs, and TOE1 depletion caused defective splicing of a pre-mRNA reporter[7]. To validate TOE1 association with the spliceosome, we performed *TOE1* knockdown followed by expression at near endogenous levels of FLAG-tagged TOE1-WT or -DE mutant in T-REx-293 cells, followed by immunoprecipitation (IP). As assessed by Northern blotting (Fig. 4a) and tandem mass spectrometry (MS) (Supplementary Fig. 7, Supplementary Table 3), TOE1 assembled with snRNAs and, with near complete overlap to previously reported IP-MS results[7], with spliceosomal proteins. The CCR4-like protein Angel2/Ccr4d, which we previously reported in complex with TOE1[4] was not detected in our IP-MS, possibly due to low cellular abundance[9].

Inspection of Northern blots for TOE1-DE-associated snRNAs revealed slower migrating U1, U2, and U5 snRNA species (Fig. 4a, asterisks). Processing of Pol-II transcribed snRNAs, U1, U2, U4 and U5, initiates with co-transcriptional cleavage by the Integrator complex downstream of the mature 3' end[10], but the mechanism mediating removal of the 3' tail to produce a mature length snRNA is unknown[11,12]. To characterize the slower migrating snRNAs, we performed 3' end sequencing of TOE1-associated snRNAs and found TOE1 bound Pol-II transcribed snRNAs that were incompletely processed at the 3' end (Fig. 4b,c, Supplementary Fig. 8 - RIP), suggesting TOE1 may mediate snRNA 3' tail processing.

TOE1-associated pre-snRNA 3' tail sequences consisted of both genome-encoded and post-transcriptionally added nucleotides, henceforth referred to as templated and untemplated tails, respectively (Figs. 4b, c, Supplementary Fig. 9a). The untemplated snRNA tails, which were previously observed by 3' end sequencing of non-coding RNA[13], were found almost exclusively on snRNAs that were either longer or shorter than annotated mature length and consisted primarily of uridines and adenosines, with TOE1-DE-associated snRNAs predominantly enriched for untemplated 3' adenosines (Supplementary Fig. 9a, b). Interestingly, 3' maturation of U2 snRNA finishes with the addition of an untemplated 3' adenosine[14], but this modification was entirely absent from TOE1-associated U2 snRNAs (Supplementary Fig. 9b). Taken together, these observations suggest that TOE1 associates specifically with pre-snRNAs that are not fully processed at the 3' end, and often, have acquired untemplated tails.

snRNAs associated with TOE1-DE contained longer tails than those associated with TOE1-WT, suggesting that TOE1 may catalytically process these tails. Consistent with this, 3' end sequencing of the total snRNA pool revealed an increased fraction of 3' end extended snRNAs upon depletion of TOE1 (Fig. 4b, Supplementary Fig. 8). Importantly, complementation with exogenous TOE1-WT rescued the snRNA 3' end defect, while TOE1-DE failed to rescue (Fig. 4b, Supplementary Fig. 8). There was little to no accumulation of 3' tails for the C/D box U3 or H/ACA box SNORA63 small nucleolar (sno)RNAs, 5.8S ribosomal (r)RNA, transfer (t)RNAs, or U6 snRNA (Supplementary Fig. 8, 10), which like Pol-II transcribed snRNAs are processed from 3' extended precursor RNAs[15–18]. These results support a catalytic role for TOE1 as a 3' to 5' exonuclease with specificity for snRNA processing.

TOE1 could promote either maturation or degradation of pre-snRNAs. To discriminate between these possibilities we tested the activity of immuno-isolated TOE1 on co-purifying pre-snRNPs *in vitro*. Northern blot analyses revealed Mg^{2+}- and active site-dependent shortening of U1, U2, and U5 pre-snRNAs by TOE1 with no overall loss of snRNA levels (Fig. 5a), suggesting a function in maturation rather than degradation. 3' end sequencing confirmed *in vitro* processing of U1, U2 and U5 by TOE1 halted at, or before, the mature 3' end (Fig. 5b, Supplementary Fig. 11a) including at an alternative U5 3' end also present in total cellular snRNA (Supplementary Fig. 8b). Consistent with a function in maturation rather than degradation, TOE1 depletion did not affect levels of unstable endogenous U1 snRNA variants (Supplementary Fig. 11b)[19], nor did it affect the levels or 3' processing of unstable U1 snRNAs mutated in the Sm binding site (Supplementary Fig. 11c, d)[20],

suggesting that the latter are targeted for degradation prior to 3' processing by TOE1. We conclude that TOE1 acts in snRNA 3' end maturation (Fig. 5c).

To determine whether TOE1 patient mutations affect snRNA processing, we performed snRNA 3' end sequencing from the patient-derived cell lines. This revealed an increased fraction of U1 and U2 snRNAs, and to a less extent, U5 snRNA, containing tails in patient fibroblasts compared to unaffected relatives, which were indistinguishable from the control (Fig. 6a, Supplementary Fig. 12a, b). Similarly, patient-derived NPC lines showed enrichment for snRNA extensions (Fig. 6b, Supplementary Fig. 12a, b). We conclude that PCH7 patient cells accumulate incompletely processed snRNAs.

Here we show patients with PCH7 harbor biallelic, damaging mutations in *TOE1*, resulting in accumulation of incompletely processed snRNAs. The involvement of TOE1 in snRNA maturation was surprising given the homology with CAF1 mRNA deadenylases and the preference of TOE1 for poly(A) RNA *in vitro*[4], but could provide a mechanistic explanation for the observed defect in pre-mRNA splicing upon TOE1 depletion[7]. The enrichment of snRNAs with untemplated 3' adenosines associated with catalytically inactive TOE1 suggests 3' adenylation as a mechanism to recruit TOE1 to pre-snRNAs (Fig. 5c). The abundance of Sm subunits in the TOE1 IP-MS analysis (Supplementary Fig. 7, Supplementary Table 3) and the absence of processing of U1 snRNA mutated in the Sm binding site (Supplementary Fig. 11d) suggest the Sm-complex as another component important for TOE1 recruitment. PARN, which has been linked to familial pulmonary fibrosis, and PARN-like DEDD deadenylases were recently found to promote 3' processing of snoRNAs and piRNAs, respectively[15,21,22], and like snRNAs, RNA targets had templated and untemplated tails upon depletion of PARN[15]. Thus, adaptation of deadenylases to function in 3' processing of non-coding RNAs in conjunction with poly(A) polymerases might be a general principle in RNA metabolism.

Our results suggest that perturbations in snRNA pools may contribute to PCH7 patient manifestations, although other currently unknown functions of TOE1 cannot be ruled out as causal to the disease. A defining feature of pontocerebellar degeneration is loss of motor neurons, akin to spinal muscular atrophy (SMA). The SMA disease gene, *Survival of Motor Neuron 1*, encodes a well-established snRNP assembly factor[23,24], and together with our results suggest that maintaining proper snRNP biogenesis is important for neuronal survival. EXOCS3, encoding a core component of the 3' RNA exonuclease exosome complex is mutated in PCH type 1[25], suggesting that there may be shared RNA targets for TOE1 and the exosome, perhaps snRNAs. Consistent with the idea that snRNAs are important for neuronal survival, a recent study identified recessive cerebellar degeneration resulting from a U2 snRNA mutation in mouse[26]. We and others recently described mutations in PCH patients in genes involved in protein synthesis, including *CLP1* and other tRNA Splicing Endonuclease (TSEN) complex components leading to misprocessing of pre-tRNAs in patient cells[27–30]. While we found no defect in tRNA 3' end processing in *TOE1* mutant cells (Supplementary Fig. 10), CLP1 has also been implicated in snRNA 3' end processing[31], suggesting a possible link between *TOE1* and *CLP1* mutations in PCH. Our data along with other recent studies[26,32] suggest defects in processing of small non-coding RNAs as a common cause of severe early onset neurodegenerative conditions.

ONLINE METHODS

Patient recruitment

Patients were enrolled and sampled according to standard local practice in approved human subjects protocols at the University of California, The Rockefeller University and The Academic Medical Center (AMC) in Amsterdam for blood, saliva and skin biopsy sampling.

Sequencing

Blood was acquired from informed, consenting individuals according to institutional guidelines, and DNA extracted using established protocols. Exome sequencing was performed on both the parents and affected member(s) from each family as previously described[34]. All variants were prioritized by allele frequency, conservation, and predicted effect on protein function, and were tested by Sanger sequencing for segregation with disease. Sequence data were analyzed with Sequencer 4.9 (Gene Codes).

Genetic mapping

Chromosomal ideogram plot were generated using the Bioconductor package *quantsmooth*[35]. Red: homozygous regions that segregate between affected and unaffected family members.

Fibroblasts culture, iPSC, NPC generation

Fibroblasts were generated from unaffected and affected dermal biopsies explants. iPSCs, neural progenitor cells were obtained as previously described[36]. Mycoplasma testing was routinely performed and all cell lines were negative.

cDNA synthesis and RT-PCR

cDNA synthesis was performed with Superscript III First-Strand cDNA synthesis system for RT-PCR (Life Technologies) and used for RT-PCR or cloning. qRT-PCR for intron-containing tRNAs were performed in triplicate on 10 ng human cDNA[37]. RT-PCR to assess *toe1* intron inclusion in *toe1* Splice blocking morphant zebrafish embryos was assessed with 10 ng zebrafish cDNA and primers flanking the first intron of zebrafish *toe1* (Supplementary Table 4).

Plasmid constructs and stable cell line generation

The open reading frame of human *TOE1* was subcloned into pcDNA5/FRT/TO-FLAG with BamHI and NotI restriction sites from pcDNA3-FLAG[38]. The open reading frame of zebrafish *toe1* or human *TOE1* was subcloned into pCS2+[39] with BamHI and XhoI. Missense mutations were generated using the QuickChange Mutagenesis Kit (Agilent). For expression of N-terminal FLAG-tagged proteins, stable HEK FLp-In T-REx-293 cell lines were generated (Invitrogen). Mycoplasma testing was routinely performed and all cell lines were negative.

Northern blotting

RNA was extracted using Trizol and separated by electrophoresis on 9% 19:1 polyacrylamide, 0.6× TBE, 8M Urea at 20 mA/gel for 2 hrs. Gels were transferred in 0.5× TBE to nylon membrane at 25V for 16 hours. Membranes were UV crosslinked, blocked with Ultrahyb Oligo (Life Technologies), and hybridized with 5' end-radiolabeled DNA oligos (see Supplemental Table 4).

3' RNA tail sequencing and analysis

RNA adapters containing bar-codes and 10–11 nucleotide randomers (see Supplemental Table 4) were ligated to the 3' ends of total and eluted RNA with T4 RNA ligase (NEB) at 16°C for 16 hr, treated with RQ1 RNase-free DNase (Promega) for 30 min at 37°C and extracted with PCA (Affymetrix). cDNA was generated using AR-17 primer with Superscript III (Life Technologies). 3' ends were amplified by Q5 DNA Polymerase (NEB) with snRNA gene-specific primers and AR-17, and then with primers D501 and D702 (Illumina; see Supplemental Table 5). For tRNA 3' end sequencing, tRNAs were gel purified from a 9% polyacrylamide urea gel and libraries were prepared by the eCLIP input RNA protocol (Days 3 and 4)[40] with AG10N and AG11N 3' RNA adapters and 3Tr3 5' DNA adapter (Supplementary Table 5). All libraries were purified on AMPure XP magnetic beads (Beckman), and validated by 2200 Tape Station (Agilent). Libraries were pooled at 4 nM, denatured with NaOH and diluted in HT1 buffer (Illumina) to 7 pM then sequenced with MiSeq Reagent Kit V2 Nano configuration (Illumina). snRNA 3' sequence reads were decomplexed, and duplicates removed based on randomer sequences. Sequences were aligned with snRNA genes (including downstream regions) from Ensembl. snRNA 3' end positions were identified based on perfect alignment with one or more snRNA genes (reads with nucleotide mismatches were removed from the analyses) and 3' untemplated tails were identified as 3' sequence additions that did not align with the genomic sequence. Each experiment generally generated 5,000–35,000 unique mapped reads with the lowest being 1,173 (Supplemental Table 4). Cumulative plots were generated from reads mapping to all U1, U4, U5, and U6 variants, whereas for U2, separate cumulative plots were generated for reads mapping to *RNU2-1* and *RNU2-2P* genes, and reads were trimmed for the A and CA tails that are post-transcriptionally added to mature RNU2-1 and RNU2-2P RNAs, respectively, to allow discrimination of the mature U2 snRNAs from those that are incompletely 3' end exonucleolytically processed (Fig. 4, Fig. 6 and Supplementary Fig. 8b and 11a).

RNA Immunoprecipitation and TOE1 activity assay

One 10 cm dish of stable cell lines expressing siRNA-resistant FLAG-*TOE1* variants was transfected twice with 20 nM siRNA (Dharmacon) targeting either luciferase (siCtrl) or the 3'UTR of *TOE1* (siTOE1) using siLentFect (Bio-Rad) 72 hrs and 24 hrs before harvest. Treatment with siTOE1 generally resulted in ~10–20% remaining protein, as determined by western blot (Supplementary Fig. 3d). At ~50% density, cells were induced for 24 hrs with tetracycline (1 ng/ml), harvested in PBS and lysed in isotonic buffer containing 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.2 mM EDTA, 0.1% Triton X-100, 1 mM PMSF, 2 μg/mL aprotinin, 2 μg/mL leupeptin, 0.1 μg/ml yeast total RNA (Roche), 80 U/mL RNaseOUT

(Invitrogen) for 10 minutes on ice. Lysates were cleared at 20K×g for 15 min at 4°C. FLAG peptide was added to 1 μg/ml and lysates were incubated for 2 hrs with Anti-FLAG M2 beads (Sigma) at 4°C, washed 8 times with NET2 (10 mM Tris, pH 7.5, 150 mM NaCl, 0.1% Triton X-100). One-fifth of the beads were eluted in protein loading buffer, and the rest was resuspended in Trizol (Life Technologies). RNA was extracted according to manufacturer's protocol. Activity assays were performed post-wash on-bead for 30 minutes at 25°C with either 2 mM EDTA or 2 mM $MgCl_2$.

Western blot

Western blots were performed with rabbit polyclonal anti-Caf1z/Toe1[4], rabbit polyclonal anti-GAPDH (Cell Signaling, 2118s), and mouse anti-Vinculin (Sigma, V9264) at 1:1000 in 5% non-fat milk in PBST. Secondary antibodies used were HRP anti-mouse and HRP anti-rabbit at 1:20,000 in 5% non-fat milk in PBST.

Animals

All animal experiments complied with Institutional Animal Care and Use Committee and occurred in a non-blinded fashion. The *Toe1* mutant mice were generated using CRISPR/SpCas9 technology. Briefly, pronuclear co-injection of 5 ng Cas9 mRNA and 2.5 ng sgRNA (targeting sequence: 5'-CTG TGT GAG ATG TTC CCA GC - 3') was performed on 143 embryos. 126 embryos were transferred into host dams for implantation resulting in 23 live pups. Sanger sequence genotyping identified only one mouse with a heterozygous single base pair frame-shifting mutation, (chr4:116806688-89insA; c.668_669insA), resulting in a *Toe1* null allele. Adult male and female zebrafish (<18 months old) from wild type (AB Tubingen) and transgenic strains were maintained under standard laboratory conditions. At least three adult pairs were used to generate embryos at 0–48 hours postfertilization (hpf) for each experiment. Translation-blocking antisense morpholino (MO) oligonucleotide (7 ng), Splice-blocking antisense MO oligonucleotide (6 ng), and Control non-targeting MO oligonucleotide (7 ng) (see Supplementary Table 5) (Gene Tools) were injected into one-cell stage embryos. Gross morphology of zebrafish was assessed at 48 hpf, and embryos were defined as affected if they had an obviously misshapen head, small eyes, and curly tail (>10% reduced head/eye size and >10° angle of tail). HuC/HuD wholemount immunostaining was used to assess the presence of neuronal tissue. For mRNA rescue experiments, 0.1 pg of *in vitro* transcribed zebrafish *toe1* or human *TOE1* mRNA (mMESSAGE mMACHINE, Ambion) was co-injected with 6 ng Splice-blocking MO into one-cell stage embryos, and assessed at 48 hpf for morphological differences and neuronal tissue (minimum of 100 embryos per condition). Immunofluorescent staining was performed as previously described[41] with rabbit polyclonal caspase-3 primary antibody (Abcam, ab13847) and with anti-HuC/HuD mouse monoclonal antibody primary (Thermo Fisher, A-21271) and Alexa 488 anti-rabbit or anti-mouse secondary antibody. Zebrafish were immobilized in agarose, and fluorescent images were acquired with a Zeiss LSM 780 and an Olympus FV1000 confocal microscope. Cells were counted using ImageJ.

Immunocytochemistry

Cells were seeded on coverslips and fixed in 4% PFA or 100% cold methanol, permeabilized with 0.1% Triton-X, and blocked with 0.1% BSA, 0.5% gelatin from cold water fish skin in

PBS. Cells were incubated with primary antibody in blocking solution overnight at 4°C, washed and incubated with secondary antibody in blocking solution and 0.4 μg Hoechst for 1 hour at room temperature. Primary antibodies used were mouse anti-Nestin (EMD Millipore, MAB5326) and rabbit anti-Pax6 (Covance, PRB-278P) at 1:1000. Secondary antibodies were Alexa 555 anti-rabbit and Alexa 488 anti-mouse at 1:500. Images were taken with an Olympus IX51 inverted fluorescent microscope.

Mass spectrometry

Liquid chromatography mass spectrometry (LC/MS-MS) assays were performed as described previously[42] from anti-FLAG IP samples from extracts of 293-T-REx cell lines stably expressing FLAG-tagged TOE1 WT or TOE1 DE at near endogenous TOE1 levels, with a parental cell line serving as a control.

Protein modeling

Needleman-Wunsch alignment of human TOE1 and CNOT7 was performed with protein-protein BLAST (pBLAST). Homologous TOE1 missense mutations, and previously published inactivating mutations (*6*), were modeled onto the predicted protein structures of human CNOT7 (PDB:2D5R) and TOE1 Zinc Finger Domain (PDB:2FC6). Patient mutations predicted to result in truncated proteins (e.g. nonsense, splice, and frameshift mutations) were excluded from analysis. Residues not in alignment were excluded (F148). PyMOL was used to create 3D renderings.

Statistics

Student's two-tailed paired t-test was employed to test the significance of accumulation of extended snRNAs as indicated (Figure 4b and Supplementary Figure 8a) (U1 snRNA: siCtrl, siTOE1 $p = 0.0054$; WT, DE $p = 0.0030$; WT, RIP WT $p = 0.00069$; DE, RIP DE $p = 0.0050$; U2 snRNA: siCtrl, siTOE1 $p = 0.014$; WT, DE $p = 0.014$; WT, RIP WT $p = 0.0058$; DE, RIP DE $p = 0.012$; U5 snRNA: siCtrl, siTOE1 $p = 0.032$; WT, DE $p = 0.014$; WT, RIP WT $p = 0.011$; DE, RIP DE $p = 0.0027$; U4 snRNA: siCtrl, siTOE1 $p = 0.033$; WT, DE $p = 0.016$; WT, RIP WT $p = 0.12$; DE, RIP DE $p = 0.042$; U6 snRNA: siCtrl, siTOE1 $p = 0.57$; WT, DE $p = 0.015$; WT, RIP WT $p = 0.021$; DE, RIP DE $p = 0.19$). Student's two-tailed unpaired t-test was employed to test the significance of Cleaved Caspase 3 positive cells in Supplementary Figure 5c (Control MO, *toe1* ATG MO $p = 0.015$)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Rea M. Lardelli[1,*], Ashleigh E. Schaffer[1,2,3,*], Veerle R.C. Eggens[4,*], Maha S. Zaki[5], Stephanie L. Grainger[6], Shashank Sathe[3], Eric L. Van Nostrand[3], Zinayida Schlachetzki[2], Basak Rosti[2], Naiara Akizu[2], Eric Scott[2], Laura Dean Heckman[2], Rasim Ozgur Rosti[2], Esra Dikoglu[2], Anne Gregor[2], Alicia Guemez-Gamboa[2], Damir Musaev[2], Rohit Mande[2], Ari Widjaja[2], Tim L. Shaw[1], Sebastian Markmiller[3], Isaac

Marin-Valencia[2], Justin H. Davies[7], Linda de Meirleir[8], Hulya Kayserili[9], Umut Altunoglu[10], Mary Louise Freckmann[11], Linda Warwick[12], David Chitayat[13,14,15], Ahmet Okay Çağlayan[16,17,18], Kaya Bilguvar[18], Huseyin Per[19], Christina Fagerberg[20], Maria Kibaek[21], Kimberley A. Aldinger[22], David Manchester[23], Naomichi Matsumoto[24], Kazuhiro Muramatsu[25], Hirotomo Saitsu[24,26], Masaaki Shiina[27], Kazuhiro Ogata[27], Nicola Foulds[28], William B. Dobyns[29], Neil Chi[30], David Traver[6], Luigina Spaccini[31], Stefania Maria Bova[32], Stacey B. Gabriel[33], Murat Gunel[17], Enza Maria Valente[34], Marie-Cecile Nassogne[35], Eric J. Bennett[1], Gene W. Yeo[3,36], Frank Baas[4,**], Jens Lykke-Andersen[1,**], and Joseph G. Gleeson[1,2,**]

Affiliations

[1]University of California San Diego, La Jolla, CA 92093, USA [2]Laboratory of Pediatric Brain Disease, Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA [3]Department of Cellular and Molecular Medicine, Stem Cell Program and Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA [4]Department of Clinical Genetics, Academic Medical Center, Meibergdreef 9,1105AZ Amsterdam, The Netherlands [5]Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo 12311, Egypt [6]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093, USA [7]Department of Paediatric Medicine, University Hospital Southampton NHS Foundation Trust, SO16 6YD, Southampton, UK [8]Pediatric Neurology and Metabolic Diseases, Universitair Ziekenhuis Brussels, Vrije Universiteit Brussel, B-1090 Brussels, Belgium [9]Medical Genetics Department, Koc University School of Medicine, Sarıyer 34450 İstanbul Turkey [10]Medical Genetics Department, İstanbul Medical Faculty, İstanbul University, Fatih 34093 İstanbul Turkey [11]Department of Clinical Genetics, The Canberra Hospital, Woden ACT 2606, Australia [12]Australian Capital Territory Genetic Service, The Canberra Hospital, Canberra City ACT 2601 Australia [13]Department of Pediatrics, Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario M5G 1X8 Canada [14]Neuroimaging, The Hospital for Sick Children, Toronto, ON, Canada [15]The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, M5G 1X5 Canada [16]Department of Medical Genetics, School of Medicine, Istanbul Bilim University, Istanbul, 34394, Turkey [17]Yale Program on Neurogenetics, Departments of Neurosurgery, Neurobiology and Genetics, Yale University School of Medicine, New Haven, CT 06510, USA [18]Department of Genetics, Yale Center for Genome Analysis, Yale University School of Medicine, New Haven, CT 06510, USA [19]Division of Pediatric Neurology, Department of Pediatrics, Erciyes University School of Medicine, Kayseri, 38039 Turkey [20]Department of Clinical Genetics, Odense University Hospital, Odense, 5000 Denmark [21]Department of Pediatric, Odense University Hospital, Odense, 5000 Denmark [22]Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington, 98101 USA [23]Department of Pediatrics, Clinical Genetics and Metabolism, University of Colorado School of Medicine, Children's Hospital Colorado, Aurora, Colorado, 80045 USA

[24]Department of Human Genetics, Yokohama City University, Graduate School of Medicine, Yokohama 232-0024, Japan [25]Gunma University School of Medicine, Department of Pediatrics, Showa-machi, Maebashi City, Gunma 371, Japan [26]Department of Biochemistry, Hamamatsu University School of Medicine, Hamamatsu 431-3192, Japan [27]Department of Biochemistry, Yokohama City University Graduate School of Medicine, Yokohama 232-0024, Japan [28]Southampton University Hospitals Trust, Southampton, Hampshire, UK [29]Seattle Children's Research Institute, Centre for Integrative Brain Research, Seattle, WA 98195, USA [30]UCSD Cardiology, University of California San Diego, La Jolla, CA 92093, USA [31]Clinical Genetic Unit, Dept. of Women, Mother and Neonates, "Vittore Buzzi" Children's Hospital, Istituti Clinici di Perfezionamento, Milan, Italy [32]Child Neurology Unit, Dept. of Pediatrics, "Vittore Buzzi" Children Hospital, Istituti Clinici di Perfezionamento, Milan, Italy [33]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA 02142, USA [34]Section of Neurosciences, Dept. of Medicine and Surgery, University of Salerno, Salerno, Italy [35]Pediatric Neurology, Université Catholique de Louvain, Cliniques Universitaires Saint-Luc, Brussels, Belguim [36]Department of Physiology, National University of Singapore and Molecular Engineering Laboratory, A*STAR, 138668 Singapore

Acknowledgments

We thank the patients and their families for their contributions to this study. Supported by the NIH R01GM077243 and R35GM118069 to J.L-A, NIH R01NS041537, R01NS048453, R01NS052455, P01HD070494, P30NS047101, the Simons Foundation Autism Research Initiative (SFARI), and the Howard Hughes Medical Institute to J.G.G, and NIH HG004659 and NS075449 and California Institute of Regenerative Medicine RB3-05009 to G.W.Y. G.W.Y. is an Alfred P. Sloan Research Fellow. E.J.B. was supported by a New Scholar award from the Ellison Medical Foundation and a Hellman Fellowship. R.M.L. is the recipient of a NRSA Postdoctoral Fellowship (NIH F32 GM106706) and is a San Diego IRACDA Fellow (NIH K12 GM06852). A.E.S. is a recipient of an A.P. Gianinni Fellowship and an NIH Pathway to Independence Award (K99HD082337). E.L.V.N. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2172-13). European Research Council Starting Grant 260888 and the Italian Ministry of Health Ricerca Corrente 2015 supported E.M.V. We thank the Broad Institute (U54HG003067 to E. Lander, UM1HG008900 to D. MacArthur), U Washington Center for Mendelian Genomics (UM1HG006493 to M. Bamshad), the Yale Center for Mendelian Disorders (U54HG006504 to R. Lifton and M.G.), and the Gregory M. Kiez and Mehmet Kutman Foundation to M.G. W.B.D. was supported by NIH R01NS050375. We thank R. Parker for U1 constructs and F. Tan for his support with protein modeling. We acknowledge M. Gerstein, S. Mane, A. B. Ekici, S. Uebe, and UCSD IGM Genetics Center for sequencing support and analysis, the Yale Biomedical High Performance Computing Center for data analysis and storage, the Yale Program on Neurogenetics, and the Yale Center for Human Genetics and Genomics.

REFERENCES

1. Goldstrohm AC, Wickens M. Multifunctional deadenylase complexes diversify mRNA control. Nat Rev Mol Cell Biol. 2008; 9:337–344. [PubMed: 18334997]
2. Anderson C, Davies JH, Lamont L, Foulds N. Early pontocerebellar hypoplasia with vanishing testes: A new syndrome? Am J Med Genet A. 2011; 155A:667–672. [PubMed: 21594990]
3. De Belle I, Wu JX, Sperandio S, Mercola D, Adamson ED. In vivo cloning and characterization of a new growth suppressor protein TOE1 as a direct target gene of Egr1. J Biol Chem. 2003; 278:14306–14312. [PubMed: 12562764]
4. Wagner E, Clement SL, Lykke-Andersen J. An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. Mol Cell Biol. 2007; 27:1686–1695. [PubMed: 17178830]
5. Namavar Y, et al. Clinical, neuroradiological and genetic findings in pontocerebellar hypoplasia. Brain. 2011; 134:143–156. [PubMed: 20952379]

6. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–249. [PubMed: 20354512]
7. Fong KW, et al. Whole-genome screening identifies proteins localized to distinct nuclear bodies. J Cell Biol. 2013; 203:149–164. [PubMed: 24127217]
8. Will CL, et al. The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. RNA. 2004; 10:929–941. [PubMed: 15146077]
9. Nagaraj N, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol. 2011; 7:548. [PubMed: 22068331]
10. Baillat D, et al. Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. Cell. 2005; 123:265–276. [PubMed: 16239144]
11. Madore SJ, Wieben ED, Pederson T. Intracellular site of U1 small nuclear RNA processing and ribonucleoprotein assembly. J Cell Biol. 1984; 98:188–192. [PubMed: 6200485]
12. Wieben ED, Nenninger JM, Pederson T. Ribonucleoprotein organization of eukaryotic RNA. XXXII. U2 small nuclear RNA precursors and their accurate 3' processing in vitro as ribonucleoprotein particles. J Mol Biol. 1985; 183:69–78. [PubMed: 2409291]
13. Goldfarb KC, Cech TR. 3' terminal diversity of MRP RNA and other human noncoding RNAs revealed by deep sequencing. BMC Mol Biol. 2013; 14:23. [PubMed: 24053768]
14. Cho HD, Tomita K, Suzuki T, Weiner AM. U2 small nuclear RNA is a substrate for the CCA-adding enzyme (tRNA nucleotidyltransferase). J Biol Chem. 2002; 277:3447–3455. [PubMed: 11700323]
15. Berndt H, et al. Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming. RNA. 2012; 18:958–972. [PubMed: 22442037]
16. Lund E, Dahlberg JE. Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNA's. Science. 1992; 255:327–330. [PubMed: 1549778]
17. Granneman S, Baserga SJ. Crosstalk in gene expression: coupling and co-regulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing. Curr Opin Cell Biol. 2005; 17:281–286. [PubMed: 15901498]
18. Phizicky EM, Hopper AK. tRNA biology charges to the front. Genes Dev. 2010; 24:1832–1860. [PubMed: 20810645]
19. O'Reilly D, et al. Differentially expressed, variant U1 snRNAs regulate gene expression in human cells. Genome Res. 2013; 23:281–291. [PubMed: 23070852]
20. Shukla S, Parker R. Quality control of assembly-defective U1 snRNAs by decapping and 5'-to-3' exonucleolytic digestion. Proc Natl Acad Sci U S A. 2014; 111:E3277–E3286. [PubMed: 25071210]
21. Izumi N, et al. Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. Cell. 2016; 164:962–973. [PubMed: 26919431]
22. Tang W, Tu S, Lee HC, Weng Z, Mello CC. The RNase PARN-1 Trims piRNA 3' Ends to Promote Transcriptome Surveillance in C. elegans. Cell. 2016; 164:974–984. [PubMed: 26919432]
23. Fischer U, Liu Q, Dreyfuss G. The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell. 1997; 90:1023–1029. [PubMed: 9323130]
24. Liu Q, Fischer U, Wang F, Dreyfuss G. The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. Cell. 1997; 90:1013–1021. [PubMed: 9323129]
25. Wan J, et al. Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. Nat Genet. 2012; 44:704–708. [PubMed: 22544365]
26. Jia Y, Mu JC, Ackerman SL. Mutation of a U2 snRNA gene causes global disruption of alternative splicing and neurodegeneration. Cell. 2012; 148:296–308. [PubMed: 22265417]
27. Schaffer AE, et al. CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. Cell. 2014; 157:651–663. [PubMed: 24766810]
28. Hanada T, et al. CLP1 links tRNA metabolism to progressive motor-neuron loss. Nature. 2013; 495:474–480. [PubMed: 23474986]

29. Budde BS, et al. tRNA splicing endonuclease mutations cause pontocerebellar hypoplasia. Nat Genet. 2008; 40:1113–1118. [PubMed: 18711368]
30. Breuss MW, et al. Autosomal-Recessive Mutations in the tRNA Splicing Endonuclease Subunit TSEN15 Cause Pontocerebellar Hypoplasia and Progressive Microcephaly. Am J Hum Genet. 2016; 99:228–235. [PubMed: 27392077]
31. Hallais M, et al. CBC-ARS2 stimulates 3'-end maturation of multiple RNA families and favors cap-proximal processing. Nat Struct Mol Biol. 2013; 20:1358–1366. [PubMed: 24270878]
32. Vance C, et al. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science. 2009; 323:1208–1211. [PubMed: 19251628]
33. Guthrie C, Patterson B. Spliceosomal snRNAs. Annu Rev Genet. 1988; 22:387–419. [PubMed: 2977088]

METHODS REFERENCES

34. Novarino G, et al. Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. Science. 2014; 343:506–511. [PubMed: 24482476]
35. Oosting J, Eilers P, Menezes R. Bioconductor package. 2012 availble at: http://www.bioconductor.org/packages/release/bioc/html/quantsmooth.html.
36. Chambers SM, et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol. 2009; 27:275–280. [PubMed: 19252484]
37. Schaffer AE, et al. CLP1 founder mutation links tRNA splicing and maturation to cerebellar development and neurodegeneration. Cell. 2014; 157:651–663. [PubMed: 24766810]
38. Wagner E, Clement SL, Lykke-Andersen J. An unconventional human Ccr4-Caf1 deadenylase complex in nuclear cajal bodies. Mol Cell Biol. 2007; 27:1686–1695. [PubMed: 17178830]
39. Turner DL, Weintraub H. Expression of achaete-scute homolog 3 in Xenopus embryos converts ectodermal cells to a neural fate. Genes Dev. 1994; 8:1434–1447. [PubMed: 7926743]
40. Van Nostrand EL, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat Methods. 2016; 13:508–514. [PubMed: 27018577]
41. Akizu N, et al. Biallelic mutations in SNX14 cause a syndromic form of cerebellar atrophy and lysosome-autophagosome dysfunction. Nat Genet. 2015; 47:528–534. [PubMed: 25848753]
42. Erickson SL, et al. Competition between Decapping Complex Formation and Ubiquitin-Mediated Proteasomal Degradation Controls Human Dcp2 Decapping Activity. Mol Cell Biol. 2015; 35:2144–2153. [PubMed: 25870104]

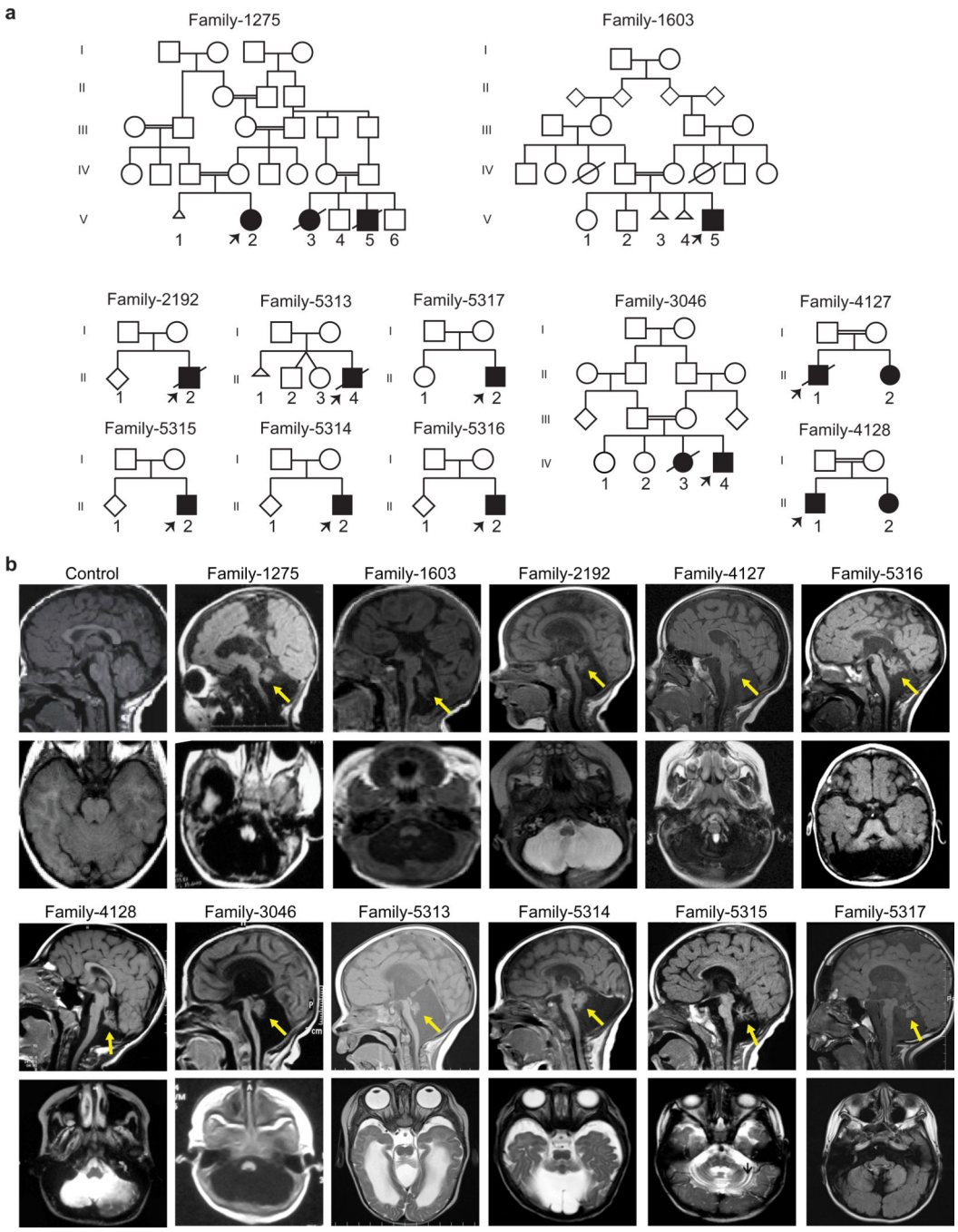

Figure 1.

TOE1 mutations lead to pontocerebellar hypoplasia with abnormal genitalia (PCH type 7). (**a**) Pedigrees of affected families showing recessive inheritance. Double bar, consanguineous marriage; open circle, unaffected female; open square, male; filled circle, affected female, filled square, affected male; triangle, spontaneous abortion; open diamond, unaffected unknown sex; diagonal line, deceased; arrow, proband. (**b**) Magnetic resonance midline sagittal (top) and axial (bottom) images showing reduced cerebellar volume (yellow arrows).

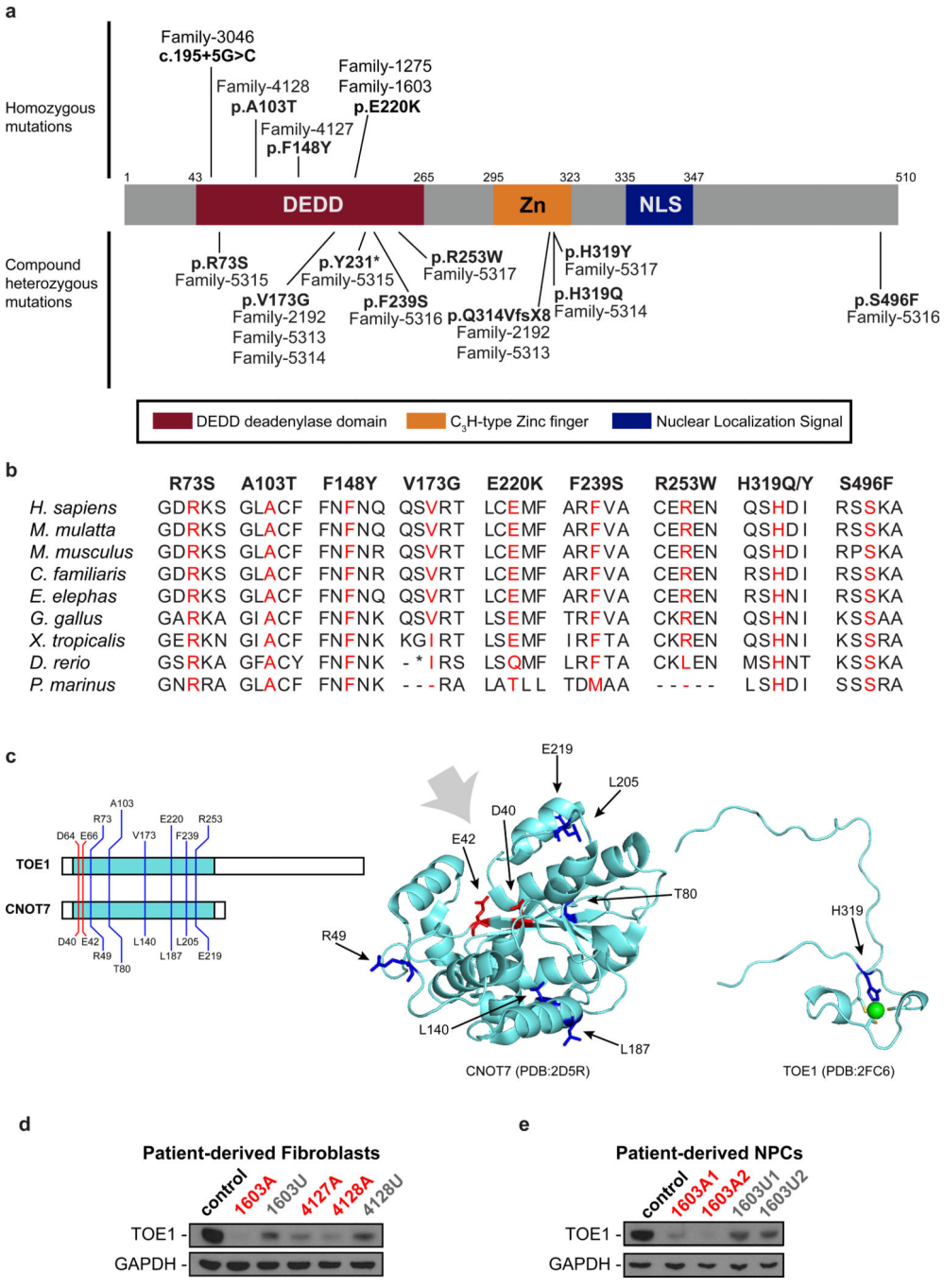

Figure 2.

TOE1 mutations occur in conserved domains and amino acid residues and reduce protein levels. (**a**) TOE1 is a 510 amino acid protein containing a DEDD deadenylase domain (red), C_3H-type Zinc finger (orange), and a nuclear localization signal (NLS, blue). Identified homozygous mutations (bold) above, compound heterozygous mutations below, with corresponding family number. (**b**) *TOE1* missense mutations showing vertebrate conservation. (**c**) TOE1 mutations (blue) modeled onto solved structures of paralog CNOT7 and Zn Finger C_3H domain of TOE1 (both teal). Residues affecting deadenylase activity

(D64, E66; red) occur in the RNA cleft (grey arrow), while the patient mutations are located on the protein surface. (**d**) Western blots (cropped) showing reduced expression of TOE1 in affected (A) compared with related unaffected (U) fibroblasts. Control: ATCC fibroblast cell line. GAPDH loading control. See Supplementary Fig. 3b for full length blots (**e**) Impaired accumulation of TOE1 protein in affected (A) compared with related unaffected (U) patient-derived neural progenitor cell (NPC) lines. GAPDH loading control.

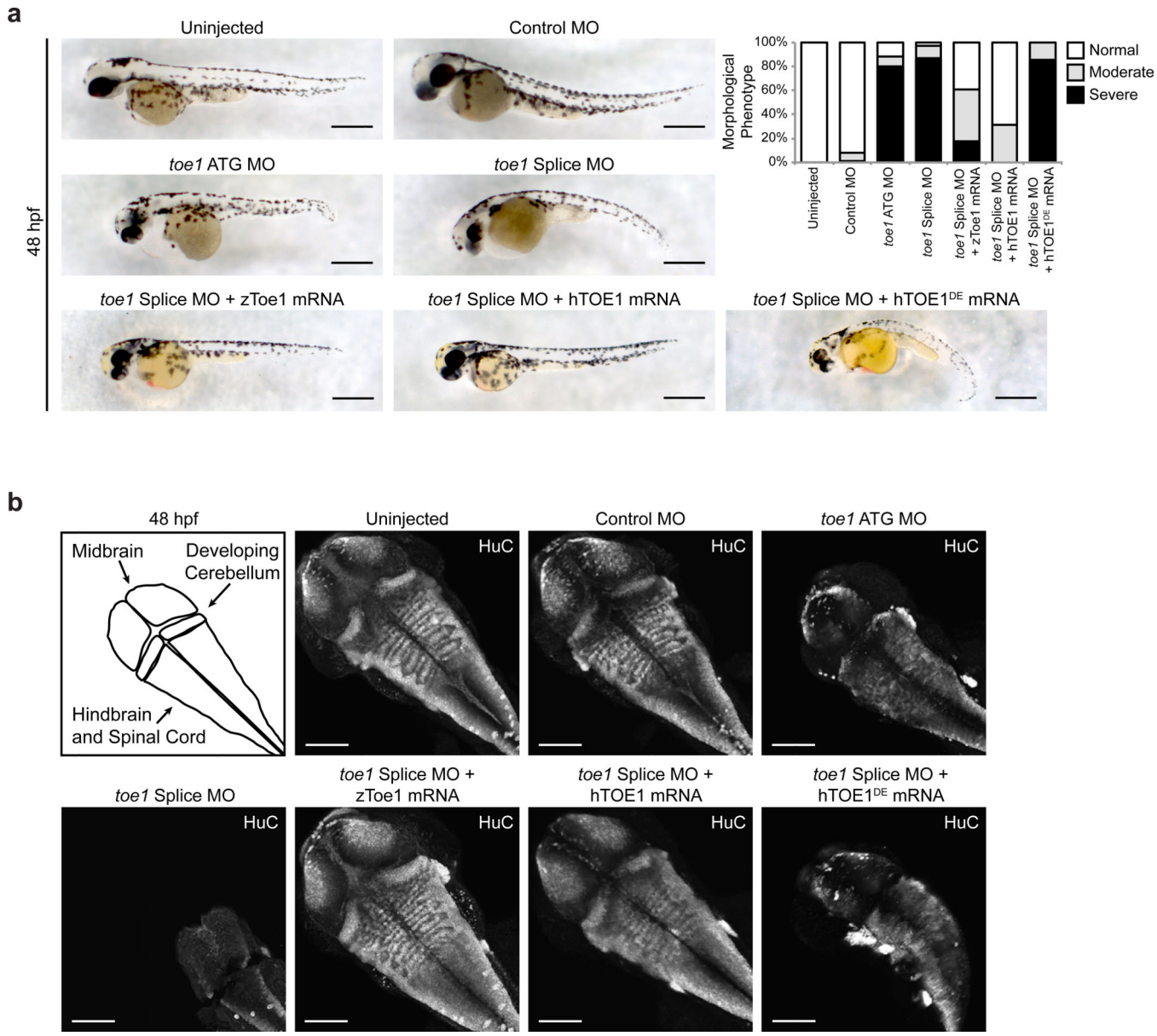

Figure 3.

Toe1 depletion in zebrafish results in structural brain defects, mimicking human PCH7 pathology. (**a**) Comparison of zebrafish at 48 hours post fertilization (hpf) injected with 6 ng Control, *toe1* ATG morpholino (MO), *toe1* Splice-blocking MO, or *toe1* Splice-blocking MO together with 0.1 pg zebrafish (z) or human (h) *TOE1* mRNAs. MO injected fish show abnormal head shape and thin, curved tails, which is rescued by addition of zebrafish or human WT, but not DE mutant *TOE1* mRNA. Scale bar = 500 μm. Quantification at bottom right. Normal: no observable phenotype; Moderate: small eyes, slight reduction in head size; Severe: small eyes, small head, thin curved tail. n=100/condition. (**b**) Maximum confocal projection of whole mount pan-neuronal HuC immunofluorescence at 48 hpf. Diagram of zebrafish anatomy (right). Midbrain, cerebellum, and hindbrain regions of *toe1* MO injected

zebrafish have reduced HuC protein levels, which were restored with WT zebrafish and human TOE1, but not human DE mutant mRNA. Scale bar = 100 μm. n = 6.

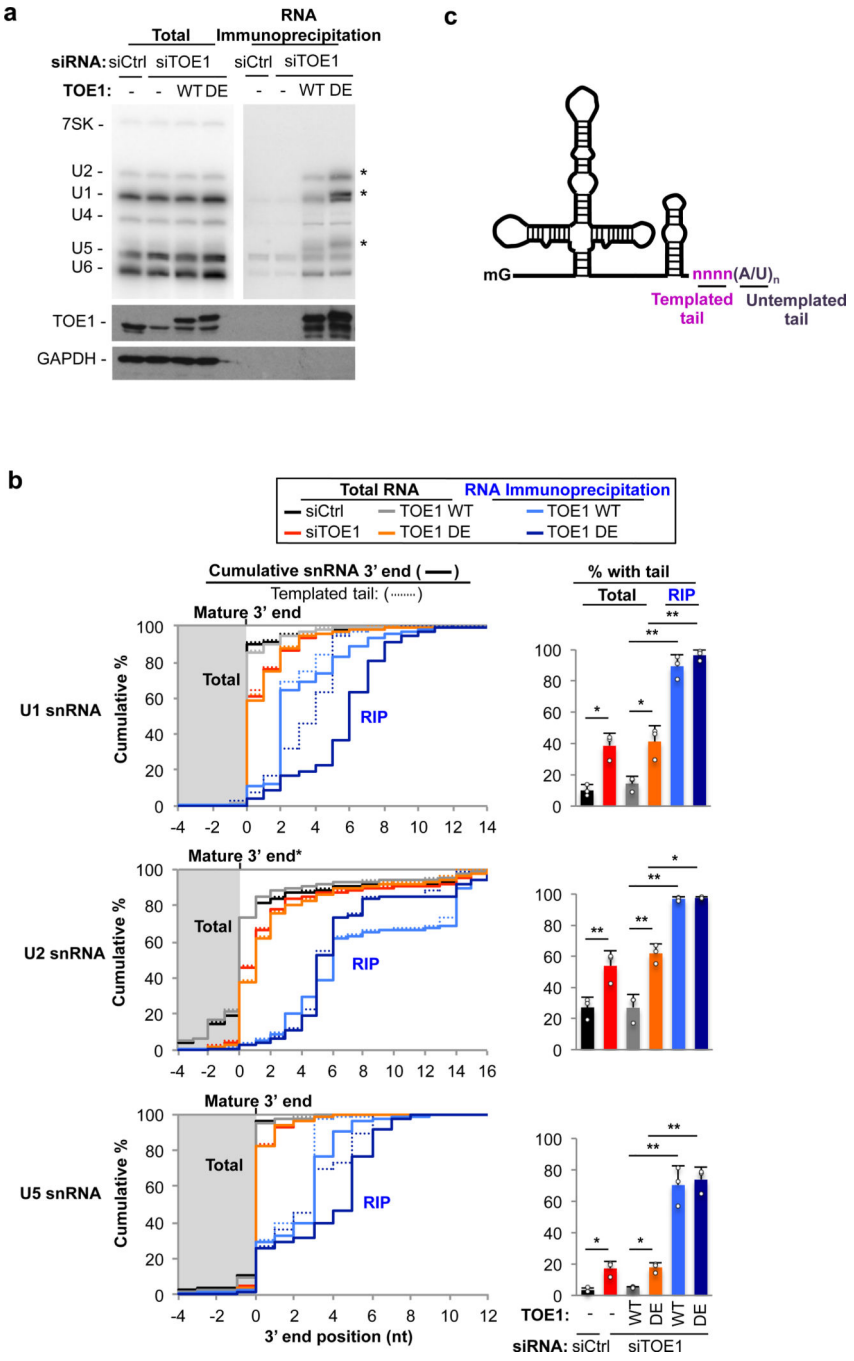

Figure 4.

TOE1 targets 3' extended pre-snRNAs. (**a**) Northern (top) and Western (middle and bottom - cropped) blots for cells treated with control (siCtrl) or TOE1 siRNA (siTOE1), then induced to express either WT or DE FLAG-TOE1 at near endogenous levels. Input samples (left). Samples from RNA immunoprecipitation with anti-FLAG (right). Arrows mark slower migrating snRNA species associated with TOE1 DE suggesting impaired processing. (**b**) Left: Cumulative plots of sequence reads for snRNA 3' ends from the average of 3 independent experiments. Position '0' refers to mature 3' end of snRNAs[33] with shaded

areas indicating 3’ end positions within the mature snRNAs. Reads terminating at position -4 or after are represented (see Supplementary Figure 8b for all reads). Dotted curves represent 3’ end positions of genome templated snRNA sequences. Solid curves mark 3’ end positions of snRNA sequences, including untemplated tails. Right: Bar graphs showing the average percent of snRNA reads with 3’ tails from 3 independent experiments. Independent experiments represented by dots. Error bars: standard deviation (S.D.) from three independent experiments and p-values (Student’s two-tailed paired t-test) *: $p<0.05$, **: $p<0.01$. Cumulative plots and bar graphs for U1 and U5 snRNAs represent reads from all snRNA variants, whereas for U2 snRNA, reads are from *RNU2-1* genes only (see Supplementary Figure 8b for *RNU2-2P* genes) and the 3’ adenosine added to mature RNU2-1 snRNAs was left out of the analysis to allow visualization of exonucleolytic processing. (**c**) Schematic of U1 snRNA processing intermediate with 3’ end templated tails (encoded nucleotides, pink) and untemplated tails (unencoded nucleotides, purple).

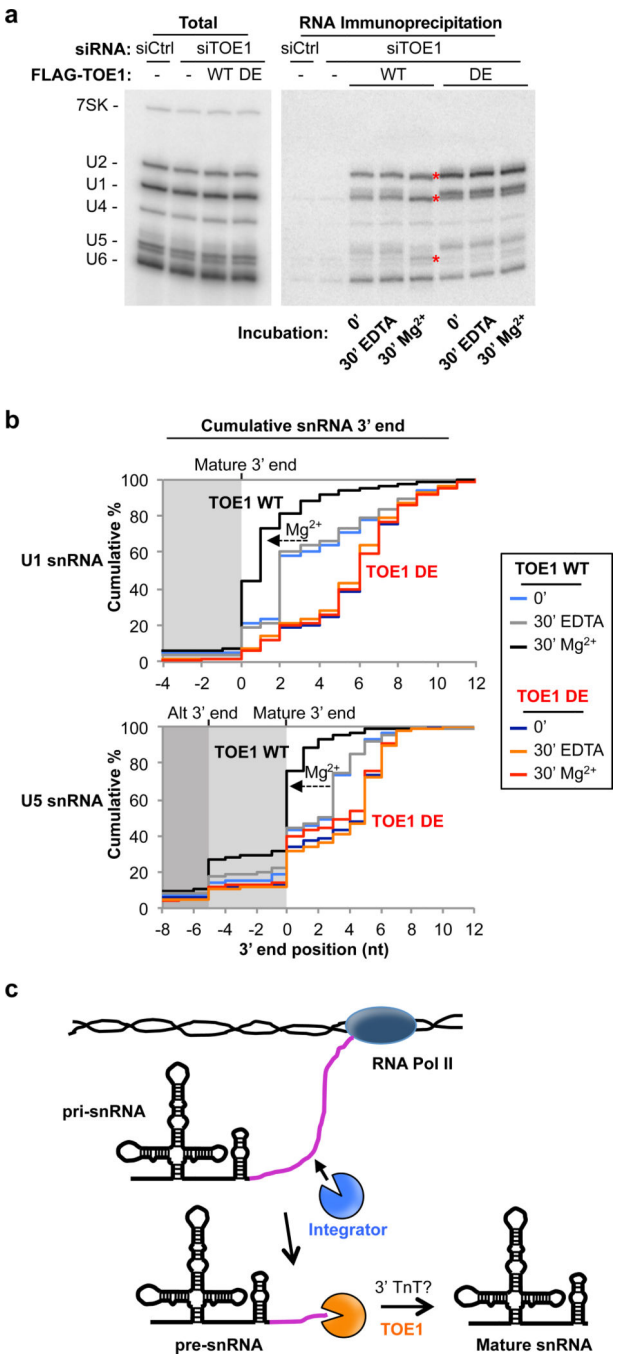

Figure 5.

TOE1 enzymatic activity processes snRNA 3' ends. (**a**) Northern blot for RNA from TOE1 activity assays. Left: input samples from RNA of cells treated with siCtrl or siTOE1, then induced to express either FLAG-TOE1-WT or -DE at near endogenous levels. Right: samples from RIP with anti-FLAG were divided and treated either with 2mM Mg^{2+} or 2mM EDTA on-bead, post-IP; asterisks mark TOE1-processed snRNA. (**b**) Cumulative plots of sequence reads for U1 and U5 snRNA 3' end positions after incubation with TOE1-WT or -DE. Dark shaded area indicates reads terminating within the mature U5 snRNA and before

the alternative 3’ end (Alt 3’ end). (**c**) Primary Pol-II transcribed snRNAs (pri-snRNAs) are co-transcriptionally cleaved by the Integrator complex. Remaining 3’ tails are subsequently processed to mature length by TOE1, in a process that might involve a 3’ terminal nucleotidyltransferase (3’ TnT).

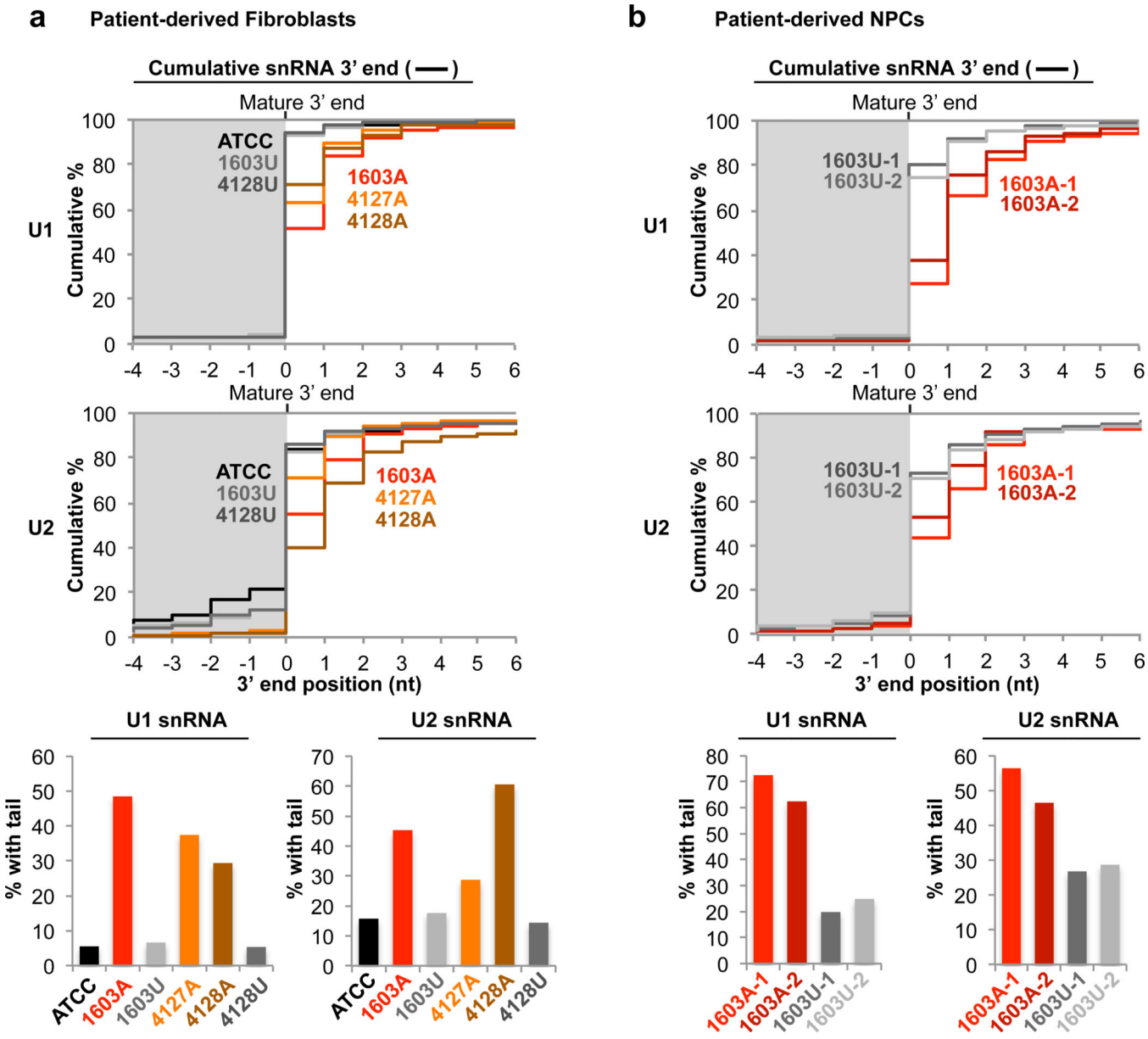

Figure 6.

Patient-derived fibroblasts and neuronal progenitor cells (NPCs) show defects in snRNA 3' ends. (**a**) Cumulative U1 and U2 snRNA 3' end sequence reads from affected (A) and unaffected (U) patient fibroblasts, showing TOE1 mutations result in snRNA tails (top). Bar graph showing percent reads with 3' tails (bottom). (**b**) Cumulative U1 and U2 snRNA 3' end sequence reads from affected and unaffected patient-derived neuronal progenitor (NPC) lines showing TOE1 mutations result in snRNA tails for NPCs (top). Percent reads with 3' end tails (bottom).